Exploring pre-service teachers' understanding of statistical variation: Implications for teaching and research

Sashi Sharma
University of Waikato, NZ
<sashi@waikato.ac.nz>

Concerns about the importance of variation in statistics education and a lack of research in this topic led to a preliminary study which explored pre-service teachers' ideas in this area. The teachers completed a written questionnaire about variation in sampling and distribution contexts. Responses were categorised in relation to a framework that identified levels of statistical thinking. The results suggest that while many of the students appeared to acknowledge variation, they were not able to provide adequate explanations. Although the pre-service teachers have had more real-life experiences involving statistics and have been involved in the study of statistical concepts at secondary school level, they still demonstrated the same misconceptions as those of younger students reported in research literature. While more students showed competence on the sampling question, they were less competent on the distribution task. This could be due to task format or contextual issues. The paper concludes by suggesting some implications for further research and teaching.

Background

Over the past 15 years, statistics has gained increased attention in our society. Many everyday activities require an understanding of statistics. Decisions concerning business, industry, employment, sports, health, law and opinion polling are made using an understanding of statistical information. Parallelling these trends, there has been a movement in many countries to include statistics at every level in the mathematics curricula. In western countries such as Australia (Australian Education Council, 1991) and New Zealand (Ministry of Education, 1992) these developments are reflected in official documents and in materials produced for teachers. Clearly the emphasis in these documents is on producing intelligent citizens who can reason with statistical ideas and make sense of statistical information.

Many statistics educators (Moore, 1997, 1990; Watson, 2007; Wild & Pfannkuch, 1999) claim that variability plays a central role in statistical think-

ing. For instance, Moore (1990) puts variability at the heart of the process of statistical thinking and describes the needs of statistical thinkers to acknowledge the omnipresence of variation and to consider appropriate ways to quantify, explain and model the variability in data. Watson (2007, p. 5) writes that, "without variation not only would the world be a very dull place but also there would be no need for statistics." According to Wild and Pfannkuch (1999), the analysis of variation is critical to the study of statistics. They identify a consideration of variation as one of the fundamental aspects of their model of statistical thinking. Wild and Pfannkuch outlined four components of variation to consider: noticing and acknowledging, measuring and modelling, explaining and dealing with, and developing investigative strategies in relation to variation. Shaughnessy and Pfannkuch (2002) believe that understanding and reducing variation are keys to success in quality management fields. Although it has been argued that variability plays a fundamental role in students' understanding and application of statistics and chance, little research attention has been given to these concepts (Ben-Zvi & Garfield, 2004; Reading, 2004; Shaughnessy, 2006, 1997; Shaughnessy, Watson, Moritz, & Reading, 1999). More research needs to be undertaken to better understand how students view and describe variation (Reading, 2004).

There is now a considerable body of international research showing how important it is for teachers to have a deep understanding of mathematics concepts and processes in order to be effective teachers of mathematics (Goulding, Rowland & Barber, 2002; Hill, Rowan & Ball, 2005; Mandeville & Lui, 1997). For instance, Hill, Rowan and Ball, suggest that teachers' discipline knowledge is linked to student achievement and improving teachers' content knowledge will improve students' performance and understanding. Within the New Zealand context, Irwin and Britt (1999) reported that the content knowledge of teachers impacted on their subsequent willingness to bring about changes to their teaching practice. Additionally, the focus on content knowledge in teacher education has become a focus in contemporary reforms. Knowledge of mathematics/statistics is a key aspect of the content knowledge referred to in the "Graduating Standards" document prepared by the NZ Teachers' Council (Wilson, 2006).

Another large body of international literature indicates that a substantial proportion of pre-service teacher education students lack confidence in their own mathematics, and in their mathematics content knowledge (Brown, McNamara, Hanley & Jones, 1999; Burgess, 2000; Goulding, Rowland & Barber, 2002; Zevenbergen, 2005). Burgess compared the probability concepts of a group of pre-service primary school teachers with the misconceptions exhibited by a group of 11 and 12-year-olds. Burgess reports that although the pre-service teachers have had more real-life experiences involving probability and have been involved in the study of probability concepts at secondary school level, they still demonstrated the same misconceptions as the younger students. Additionally, whatever probability and statistics knowledge teachers have acquired at secondary school or university was not usually taught in a way designed to develop understanding or correct intuitions. Reading (2004) states that data reduction learning experiences for secondary

school students mostly deal with measures of central tendency and hence few of these students bother with measures of variation. Reading adds that teachers find the study of measures of variation particularly cumbersome and hence have difficulty developing these concepts with students or leave it completely.

Concerns about the importance of variation in statistical thinking and a lack of research in this area determined the focus of this study. Overall, the study was designed to investigate pre-service teachers' acknowledgement of variation in sampling and distribution environments.

Research on statistical variation

To illustrate the undue confidence that people put in the reliability of small samples, Tversky and Kahneman (1974) gave the following problem to tertiary students.

> Assume that the chance of having a boy or girl baby is the same. Over the course of a year, in which type of hospital would you expect there to be more days on which at least 60% of the babies born were boys?
> (a) In a large hospital
> (b) In a small hospital
> (c) It makes no difference

Most subjects in Tversky and Kahneman's study (1974) judged the probability of obtaining more than 60% boys to be the same in the small and in the large hospital. According to Tversky and Kahneman (1974) the representativeness heuristic underlies this misconception. People who rely on the representative heuristic tend to estimate the likelihood of events by neglecting the sample size or by placing undue confidence in the reliability of small samples. However, the sampling theory entails that the expected number of days on which more than 60% of the babies are boys is much more likely to occur in a small hospital because a large sample is less likely to stray from 50%.The concept of sample size is an important idea in statistical decision making.

On the other hand, rather than neglecting the effect of sample size, Shaughnessy (1997) provides evidence that students may actually superimpose a sampling setting on a question where none is there to begin with, in order to establish a centre from which to predict. For instance, consider the following task given to a sample of tertiary students at the beginning of a class in statistics:

> A fair coin is flipped 5 times in succession. Which do you feel is more likely to occur for the five flips? Why?
> (A) HTTHT
> (B) HHHHH
> (C) They have the same chance of happening.

The explanations provided by the participants indicated a great variety of conceptions, and interpretations of the problem. For instance, the following explanations came from respondents choosing option (A):

"I would go with (A) only because it more closely approximates the ratio of 50–50, but in such a small sample anything is possible."

"The chances are 50/50 no matter what, and sequence A is more likely."

"(A) because I think it would be more likely to have a series of two of the same than to have five of the same."

"I would say both are equally likely on any particular instance, although the long term results would gravitate to a result more like A." Shaughnessy (1997, p. 7)

The notion of a representative sample that is so helpful in the Tversky and Kahneman (1974) question can cause problems when applied in the above context. There is no sample in the above question, there is just the sample space. Some of the respondents appeared to superimpose a sampling context on the original question in order to employ the representativeness strategy in their responses when they said, "In a small sample anything is possible and long term results would gravitate to a result more like A."

In recent years, concern over a lack of attention to variation has prompted researchers to explore students' understanding of this concept in more depth (Reading, 2004). Shaughnessy et al. (1999) surveyed 324 students in grades 4–6, 9 and 12 in Australia and the United States using a variation of an item on the National Assessment of Educational Progress (Shaughnessy & Zawojewski, 1999). Three different versions (range, choice, list) of the following task were presented in a *before* and in a *before and after* setting. In the latter setting students did the task both before and after carrying out a simulation of the task.

> A bowl has 100 wrapped hard lollies in it. Twenty are yellow, 50 are red, and 30 are blue. They are well mixed up in the bowl.
> Jenny pulls out a handful of 10 lollies, counts the number of reds, and tells her teacher. The teacher writes the number of red candies on a list. Then, Jenny puts the lollies back into the bowl, and mixes them all up again.
> Four of Jenny's classmates, Jack, Julie, Jason, and Jerry do the same thing. They each pick ten lollies, count the reds, and the teacher writes down the number of reds. Then they put the lollies back and mix them up again each time.

Responses were categorised according to their centre and spreads. While there was a steady improvement across grades on the centre criteria, there was no clear corresponding improvement on the spread criteria. There was considerable improvement on the task among the students who repeated it after the simulation. The researchers suggested that the lack of clear growth on spreads and variability and the inability of many students to integrate the two concepts (centres and variation) on the task may be due to instructional neglect of variability concepts.

As part of a larger study (Sharma, 1997), I used the following item to explore high school students' understanding of sampling variation.

Shelly is going to flip a coin 50 times and record the percentage of heads she gets. Her friend Anita is going to flip a coin 10 times and record the percentage of heads she gets.

Which person is more likely to get 80% or more heads? Explain your answer.

The students were interviewed by myself and interviews were tape recorded and transcribed for analysis. From a statistical point of view, more than 80% heads is more likely to occur in the small sample because the large sample is less likely to stray from 50%. However, none of the students' responses were considered statistical on this item, students based their reasoning on their cultural beliefs such as luck, everyday experiences and intuitive strategies such as equiprobability bias (Lecoutre, 1992). Students who used this bias tended to assume that random events were equiprobable by nature. The students responded that both Shelly and Anita will get the same number of heads because heads and tails were equally likely. Even repeated probing by myself did not induce any consideration for sample size. Two students even altered their data to this problem to align it with their personal preferences when they said they should be tossing it the same number of times.

Watson and Kelly (2003) considered students predictions and explanations for outcomes when a normal six-sided die is tossed 60 times. Since the task was part of a larger study, they were able to consider differences across grades 3 to 9 students' change in performance after some classroom chance and data experiences that were devised to enhance appreciation of variation. The researchers used a five code hierarchy to analyse the responses: pre-structural, uni-structural, transitional, multi-structural and relational. The students using the relational level responses used appropriate variation and explanations reflecting the random nature of the process. Only 7% of students across grades 5 and 7 responded appropriately. A decrease was evident in grade 9. The researchers suggest that teachers themselves may be a useful focus of research in terms of their own understanding of expectation and variation. In the current study, two open-ended questions were used to determine specific student conceptions about variation and the factors that contribute to these constructs. An overview of the research design follows, after which I will discuss the results of my study.

Overview of the study

The research setting was a graduate mathematics education course situated in the second semester for prospective primary teachers at a large university. A group of 24 pre-service teacher education students completed a questionnaire about variation during one of the tutorials. All these students were in their final year of education. After completing the questions, students were asked to choose whether or not the information they had provided could be included in a research project designed to explore ways to strengthen the mathematical understanding of pre-service teacher education students. They were asked to sign their names to indicate consent.

The tasks were selected and adapted from those used by other researchers. The birth problem (Item 1) attempted to explore students' understanding of variation in a sampling setting. The students had to select the appropriate option and provide their reasoning. The die question (Item 2) was used to elicit students' ideas about expected variation embedded in distributions of experimental outcomes. Students generated their own distributions. Responses demanded both numerical and qualitative descriptions. In both these questions, the students had to consider measures of variation to explain their reasoning, hence this is the central notion to which I refer in both items.

Item 1

Half of all newborns are girls and half are boys. Hospital A records an average of fifty births a day. Hospital B records an average of ten births a day. On a particular day, which hospital is more likely to record 80 percent or more female births?

(a) Hospital A (with fifty births a day)

(b) Hospital B (with ten births a day)

(c) The two hospitals are equally likely to record such an event.

Please explain your answer.

Item 2

(a) Imagine you threw a die 60 times. Fill in the table below to show how many times each number might come up.

Number on die	How many times it might come up?
1	
2	
3	
4	
5	
6	
TOTAL	60

(b) Why do you think these numbers are reasonable?

Analysis

Students' responses to Item 1 were categorised both on the basis of their appreciation (option (b)) and non-consideration (option (c)) for variation. It must be noted that none of the students chose option (a). Students' numerical responses on Item 2 were coded on two scales, a centring scale (10, 10. 10. 10. 10. 10) and a scale for variation (low, appropriate, high). The criteria for determining the appropriateness of variation displayed in the numerical answers was the same as that of Watson and Kelly (2003). Appropriate variation was demonstrated if the standard deviation in the responses fell between 1.2 and 4.7. Since the teachers used a variety of reasoning to justify their

predictions over the two items, I created a simple three category framework that could be helpful for describing research results and planning teaching. The three categories in the model are: non-statistical, partial-statistical and statistical. The term statistical is used in this paper for the appropriate responses. However, I am aware that such a term is not an ideal one. Student possess interpretations and representations which may be situation specific and hence these ideas have to be considered in their own right. Statistical simply means what is usually accepted in standard statistics text-books. Students using the partial statistical responses indicated some consideration of variation in their predictions/choices. However, their justifications indicated either little or no consideration for variation, Two levels of responses were identified within this category. More details of the categories and levels are given in Table 1.

Table 1. Description of categories of responses.

Response type	Description	Examples
Statistical	Appropriate variation and explanation	Item 1: b. because the higher the sample size, the more likely you will be towards having an average of 50%, hence the lower the sample size the more likely to be 80%. Item 2: 12, 11, 9, 10, 8. 9 — because they are around the expected but you can't really tell.
Partial-statistical	There were two types of partial-statistical responses. One type (Level 2) realised some conflict of probability theory and variation in their predictions. However, the explanations did not reflect any consideration for variation. The other type (Level 1) produced responses based on the equiprobability bias or made calculation errors.	Item 1 Level 2: b. because there are less babies born so less have to be male. Level 1: c. because each hospital has the same chance of reaching 80% as its 50/50 per child being born. Item 2 Level 2: 6, 10, 11, 9, 9, 15 because I am lucky that way. Level 1: 10, 10, 10, 10, 10, 10 because each number has the same chance of being rolled. 10, 10, 10, 10, 10, 10 because 6/60 0r 1/10 of a chance of each number being rolled.
Non-statistical	Students who gave non-statistical responses did not demonstrate any variation. They predicted numbers that did not sum to 60. The explanations were based on idiosyncratic phrases or there were no explanations.	Item 1: c. because the chance is the important factor not the number of births. Item 2: 10, 10, 10, 10, 10, 10 because it is a random process 60, 60, 60, 60, 60, 60 because each number has equal chance of being thrown.

Results

In this section the types of responses are summarised and the ways in which the students have explained their thinking is described. Typical responses are used for illustrative purposes. Throughout the discussion, Sn is used for the *n*th student.

Table 2. Response types for the two items (n = 24).

	Number of students using the response		
Response type	**Baby problem**	**Die problem**	**[Both problems]**
Non-statistical	2 (S21, S24,)	4 (S8, S11, S12, S24)	[1]
Partial-statistical	15 (S1, S3, S4, S5, S8, S9, S10, S11, S12, S13, S14, S15, S17, S20, S23,)	18 (S1, S2, S3, S4, S5, S6, S7, S9, S10,, S13, S14, S15, S17, S18, S20, S21, S22, S23)	[4]
Statistical	7 (S2, S6, S7, S16, S18, S19, S22)	2 (S16, S19)	[2]

Statistical responses

From a statistical point of view, more than 80% of female births is more likely to occur in Hospital B because the large sample is less likely to deviate from 50% (Item 1). To be considered statistical on Item 2, responses had to display appropriate variation and also provide explanations reflecting the random nature of the process. Table 2 shows that while seven students managed to respond in a statistical manner on Item 1, only two did so on Item 2. The following responses come from this category:

> b. Because 10 births a day is not a sufficient number to produce a reliable result.
> Because the sample is smaller it has more variability. (S16)
> b. Short frequencies are more likely to deviate from the true probability. (S22)
> Because each number should come up roughly 10 times, give or take a few. The more times the dice is thrown the better. (S16)

Partial-statistical responses

Of the 15 students with partial-statistical responses on Item 1, seven used level 2 type of responses whereas the rest based their reasoning on intuitions such as equiprobability bias (Level 1). One of the keys to understanding variability is balancing the ideas of theoretical and experimental probability. Students who based their explanations on the equiprobability bias tended to assume random events to be equiprobable by nature. They were unable to integrate expectation and variation in their responses. Students producing Level 2

types of responses acknowledged variation in their predictions. However, the explanations did not indicate any consideration for variation. The following are indicative of level 2 and level 1 types of responses respectively.

> b. Because only 8 of 10 have to be girls. In (a) 40 of the 50 have to be girls. (L2, S11)
> b. Because the sample is smaller so 8/10 is more likely than 40/50 girls. (L2, S23)
> c. There is always a chance that both hospitals might record 80% female births because probability is to do with equally likely outcomes. (L1, S17)

Of the 18 students whose responses have been classified as partial-statistical on Item 2, 15 responded with no variation in their predictions and based their reasoning on equal probability or were part-way to providing an appropriate explanation but needed more detail and precision. These responses are equivalent to Level 1 type of responses (Table 1).

> 10, 10, 10, 10, 10, 10. Because each number has one in six chance of being thrown. (S10)
> There are 6 numbers and they all have an equal chance of coming up ie 60/6=10 each. (S20)
> Because assuming the die is weighted evenly you are equally likely to throw either number. The sample is big enough to make it reasonable to assume an even chance (S21)

Three students provided Level 2 type of responses. Although the students responded with reasonable variation, they did not provide adequate explanations. The following explanations are indicative of this level.

> 9, 10, 10, 11, 12, 8. Because it is unlikely each number will come up an equal number of times, even though the probability is 6/60 for each number. (S23)
> 8, 10, 12, 16, 7, 7. Because any set of numbers is possible as long as they sum to 60. (S4)

Non-statistical responses

Two students judged that the probability of obtaining more than 80% of females was the same for both hospitals because the chance was the important factor not the number of births. Thus the base rate data of 80% variability was completely ignored because it did not have any implications. The four students with responses in this category for Item 2 used the centre criteria for prediction. Three of these students did not give any explanations or used terms such as random for their predictions whereas one appeared to have adapted the expectation rule. The student suggested that there was 6/60 or 1/10 of a chance of each number being rolled. The response reveals the resilience of school learnt rules and procedures (without understanding) and the impact these have on student thinking and learning.

Discussion

The thinking of most of the pre-service teachers in this survey was heavily influenced by equally likely and expectation conceptions rather than a consideration of variability. Although some students do appear to possess notions of variability, they were often unable to integrate expectation and variability into their explanations. After discussing the findings in a broader context, this section suggests some implications for further research.

Sampling variability: A broader context

The survey results indicate that variability concepts of pre-service teachers are not significantly more sophisticated than those of younger students. The findings are those of the Burgess (2000) and Watson and Kelly (2003) studies. For instance, in the Watson and Kelly study (2003), 7% of students across grades 5 and 7 responded appropriately to Item 2. In the present survey, 8% of the teachers responded appropriately. This indicates that textbook-type of exercises to do with theoretical probability are insufficient to help students develop a complete understanding of chance events. I agree with Watson and Kelly in recommending that more explicit and repeated recognition of both variation and expectation is needed if a genuine appreciation of variation is to be achieved.

According to Tversky and Kahneman (1974) and Shaughnessy (1997), the representativeness strategy underlies the sample variability misconception. The results of this survey provide evidence that most students did not rely on the representativeness strategy but based their thinking on the equiprobability bias. One explanation for this could be classroom emphasis on classicist probabilities rather than frequentist approach. Students appreciate equally likely outcomes but fail to conceptualise the variation that can emerge across a number of repetitions of the event. In short, they are unable to integrate expectation and variation (uncertainty) into the sampling construct. Another reason could be the wording of the research questions. For instance, the word “fair” in Shaughnessy’s study (1997) indicates a purposeful construction of the situation — a word that is missing from Item 2 and students may have responded differently to these situations. Unlike my study (Sharma, 1997), none of the students in this survey based their reasoning on everyday experiences and beliefs such as outcomes can be controlled. One possible explanation for this could be that the contexts and format for the tasks were quite different and the students were different ages with different statistical and cultural backgrounds, hence the reasoning employed was different.

The results show that students did not explicitly use words dealing with variation (spread, deviation). These findings are similar to those reported by Watson and Moritz (2000) and Shaughnessy et al. (1999). Moreover, many teachers gave answers that were partially correct but did not contain enough detail and did not say precisely what they meant. This is a matter of concern because this will hinder student teachers’ ability to facilitate classroom discourse and conceptual understanding.

The results suggest that some fundamental thinking was absent from some

responses. For instance, there appears to be a lack of conceptual understanding in the response of the pre-service teacher who said there was 6/60 or 1/10 of a chance of each number being rolled (Item 2). Their response did not consider the constraint that the six predicted values must add to 60. The finding is consistent with the findings of Bakker (2004) and Zevenbergen (2005). Some pre-service teachers offered responses that had calculation or interpretation errors. This is highly problematic because this can hinder the students' capacity to stimulate class discussion and to identify errors offered by students.

This preliminary survey was just a first phase towards exploring pre-service teachers' conceptions of variability. It suffers from all limitations that accompany a written questionnaire. Moreover, the open-ended nature of the questions and the lack of guidance given to students regarding what was required of them certainly influenced how students explained their understanding. Some of the conceptions addressed in this paper may actually be due to misinterpretation of the questions. Given the subtleties of interpretations, it is unlikely that the items used in the survey would have discriminated finely enough. For instance, although Item 1 was intended to be about variation, students may have interpreted it as if it were a question involving expectation. Some students clearly had difficulty explaining their thinking. Although the study provides some valuable insights into the kind of thinking that students use, the conclusions cannot claim generality because of a small sample. Some implications for future research are suggested by the findings despite the limitations of this study.

Implications for teaching and research

The results show that many teachers were unable to integrate centres and variation, they relied on expectation in their explanation. Since these teachers are adult products of secondary schooling, this issue needs to be addressed in high school mathematics courses to ensure that students understand the important role that variability plays in statistical reasoning.

One implication for further research could be to replicate the present study and include a larger sample of students from different educational backgrounds to claim generality. Probably there is a need to conduct individual interviews with students in order to probe their conceptions of variability at a greater depth. A sample of these students could also be interviewed while they gather actual data on the die question (Item 2) to see if the variation in results of trials influences their predictions.

Another implication relates to contexts. The picture of students' thinking in regards to variation is somehow limited because students responded to only two items. There is a need to include more items using different chance contexts such as drawing objects from containers and spinning spinners in order to explore students' conceptions of variation and related contexts in much more depth. It is also important for future research to employ a variety of task formats. Perhaps extending the question to include range and choice versions (Shaughnessy et al., 1999) and Green's (1983) graphical representation might be used. It could be useful to ask students to respond to what they

think they might get if the die rolling was repeated a second time. The survey results show that task context can bring in multiple interpretations and possibly different kinds of abstractions. For instance, while seven students managed to respond in a statistical way on Item 1, only two responses were considered appropriate on Item 2. At this point it is not clear how a learner's understanding of the context contributes to his/her interpretation of data. Research on what makes this translation difficult for students is needed.

Third, this small scale investigation into identifying and describing students' reasoning has opened up possibilities to do further research at a macro-level on students' thinking and to develop more explicit categories for each level of the framework. Such research would validate the framework of response levels described in the current study and raise more awareness of the levels of thinking that need to be considered when planning instruction and developing students' statistical thinking.

Many pre-service teachers gave partial explanations, but needed more detail or precision. Tutors need to assist pre-service teachers to express what they already know with more precise mathematical language. In the course of discussions, comparison of several answers may be made. This might lead to judgements about what might constitute a good explanation and draw attention to missing details. These implications parallel those described by Ministry of Education (1992); communicating mathematically is considered an essential skill in the New Zealand mathematics curriculum document (Ministry of Education, 1992) which has an entire sub-strand devoted to this aspect of mathematics.

Finally, like the tertiary students in this study, secondary school students are likely to resort to partial-statistical or non-statistical explanations. Research efforts at the secondary level are crucial in order to better understand how students view variation and to inform teachers and curriculum writers. Teachers need to provide opportunities for students to integrate their understanding of expectation and variation. Responses to sampling tasks such as those used in research could provide starting points for comparing estimated and experimental outcomes.

References

Australian Education Council (1991). *A National Statement on Mathematics for Australian Schools.* Carlton: Curriculum Corporation.

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64–83.

Ben-Zvi , D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal, 3*(2), 4–6.

Ben-Zvi D. & Garfield, J. (2004). Research on reasoning about variability: A forward. *Statistics Education Research Journal, 3*(2), 42–61.

Brown, T., McNamara, O, Hanley, U. & Jones, J. (1999). Primary student teachers' understanding of mathematics and its teaching. *British Educational Research Journal, 25*(3), 299–322.

Burgess, T. (2000). Are teachers' concepts more sophisticated than those of their students? In J. Bana, & A. Chapman (Eds), *Mathematics Education Beyond 2000, Vol. 1* (Proceedings of the 23d Annual Conference of the Mathematics Education Research Group of Australasia, pp. 126–133). Sydney: MERGA.

Green, D. R. (1983). School pupil's probability concepts. *Teaching Statistics, 5*(2), 34–42.

Goulding, M., Rowland, T. & Barber, P. (2002). Does it matter? Primary teacher trainees' subject knowledge in mathematics. *British Educational Research Journal, 28*(5), 689–704.

Hill, H., Rowan, B. & Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.

Irwin, K. & Brit, M. (1999). Teachers' knowledge of mathematics and reflective professional development. In B. Jaworski; T. Wood, & S. Dawson (Ed.), *Mathematics Teacher Education: Critical International Perspectives* (pp. 91–101). London: Falmer Press.

Lecoutre, M. (1992). Cognitive models and problem spaces in purely random situations. *Educational Studies in Mathematics 23*, 557–568.

Mandeville, G. K. & Lui, Q. (1997). The effect of teacher certification and task level on mathematics achievement. *Teaching and Teacher Education, 13*(4), 397–407.

Ministry of Education (1992). *Mathematics in the New Zealand Curriculum*. Wellington: Author.

Moore, D. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123–165.

Moore, D. S. (1990). Uncertainity. In L. Steen (Ed.), *On the Shoulders of Giants: New Approaches to Numeracy* (pp. 95–137). Washington, DC: National Academy Press.

Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal, 3*(2), 84–105.

Sharma, S. (1997). *Statistical ideas of high school students: Some findings from Fiji*. Unpublished doctoral thesis. Waikato University, Hamilton, New Zealand.

Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. In G. Burrill (Ed.), *Thinking and Reasoning with Data and Chance* (pp. 77–78). Reston: The National Council of Teachers of Mathematics.

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph & K. Carr (Eds), *People in Mathematics Education. Mathematics Education Beyond 2000* (Proceedings of the 23d Annual Conference of the Mathematics Education Research Group of Australasia, pp. 6–22). Hamilton: MERGA.

Shaughnessy, J. & Pfannkuch, M. (2004). How faithful is old faithful? Statistical thinking: A story of variation and prediction. *The Mathematics Teacher, 95*(4), 252–259.

Shaughnessy, J. M., Watson, J. Moritiz, J. & Reading, C. (1999). *School mathematics students' acknowledgement of statistical variation*. Paper presented at the 77th Annual National Council of Teachers of Mathematics Conference, San Francisco, CA.

Shaughnessy, J. M. & Zawojewski, J. S. (1999). Secondary students' performance on data and chance in the 1996 NAEP. *The Mathematics Teacher, 92*(8), 713–718.

Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Watson, J. M. (2007). The foundations of chance and data. *Australian Primary Mathematics Classroom, 12*(1), 4–7.

Watson, J. M. & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*(1), 44–70.

Watson, J. M. & Kelly, B. A. (2003). Predicting dice outcomes: The dilemma of expectation versus variation. In L. Bragg, C. Campbell & G. Herbert (Eds), *Mathematics Education Research: Innovation, Networking, Opportunity* (Proceedings of the 26rd Annual Conference of the Mathematics Education Research Group of Australasia, pp. 728–735). Melbourne: MERGA.

Wild, C. J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–265.

Wilson, K. (2006). Standards for graduating teachers: Development in Aotearoa New Zealand. In *Proceedings of the 2006 Australian Teacher Education Association Conference*.

Zevenbergen, R. (2005). Primary preservice teachers' understandings of volume: The impact of course and practicum experiences. *Mathematics Education Research Journal, 17*(1), 3–25.